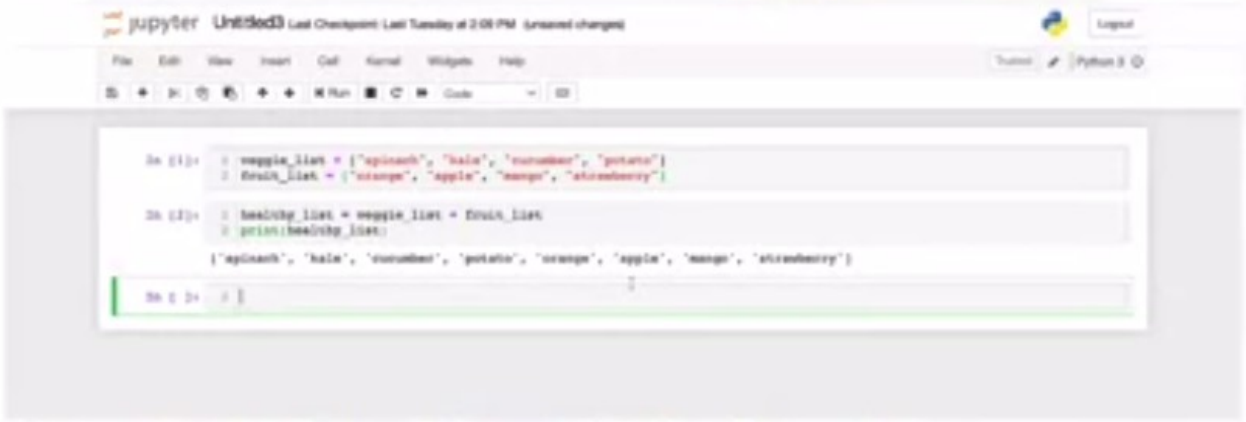
Jupiter Notebook

What is a Jupyter Notebook

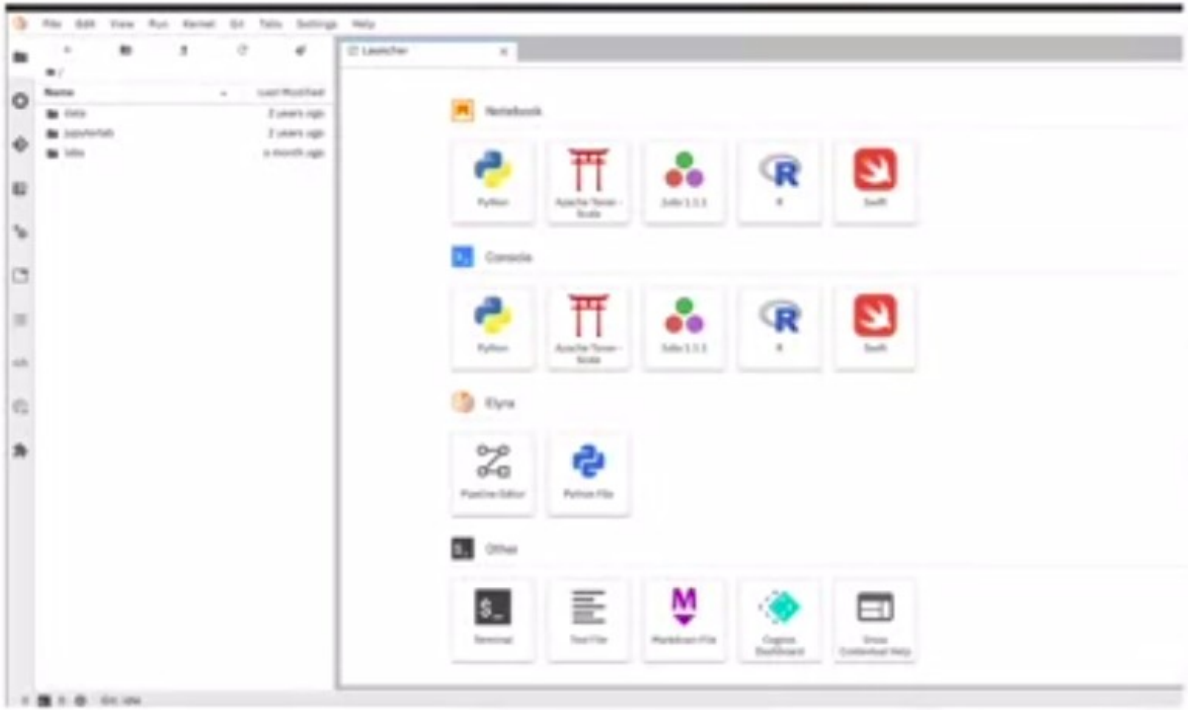


- Jupyter Notebook is a tool for recording Data Science experiments
- It allows a Data Scientist to combine text and code block in a single file
- It generates plots and tables within the file
- Notebooks can be exported as pdf and html files

IBM Developer

SKILLS NETWORK

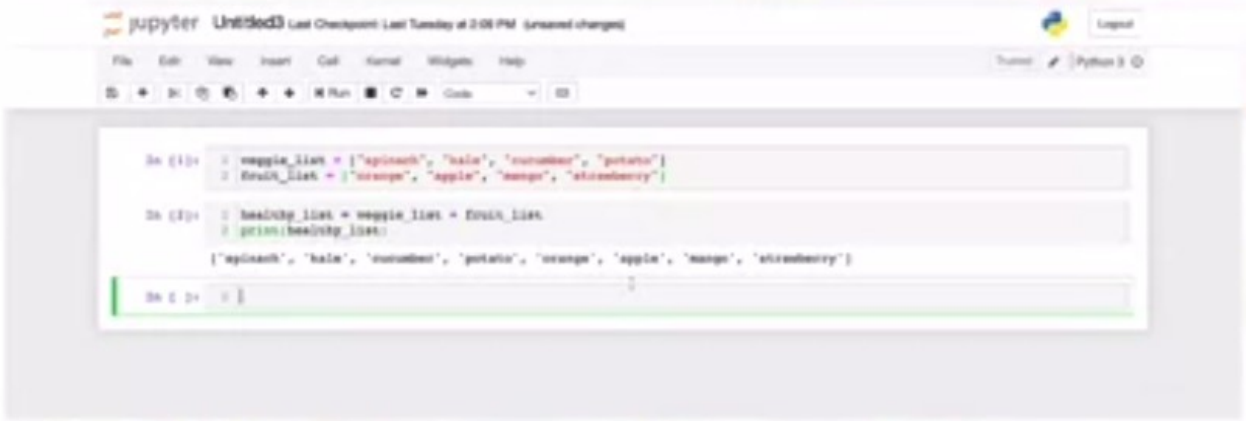
JupyterLab



JupyterLab:

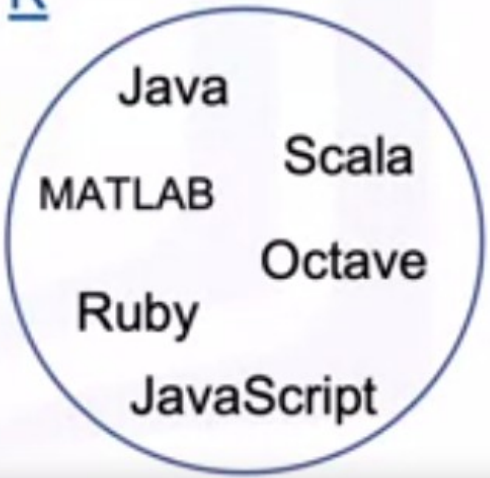
- is an interactive environment for Jupyter Notebooks,
- allows for real time editing,
- is compatible with several file formats
- and is open source

What is a Jupyter Notebook



Jupyter stands for:

- Julia
- Python
- R



- Jupyter Notebook is a tool for recording Data Science experiments
- It allows a Data Scientist to combine text and code block in a single file
- It generates plots and tables within the file
- Notebooks can be exported as pdf and html files

IBM Developer

SKILLS NETWORK

What is a Kernel?

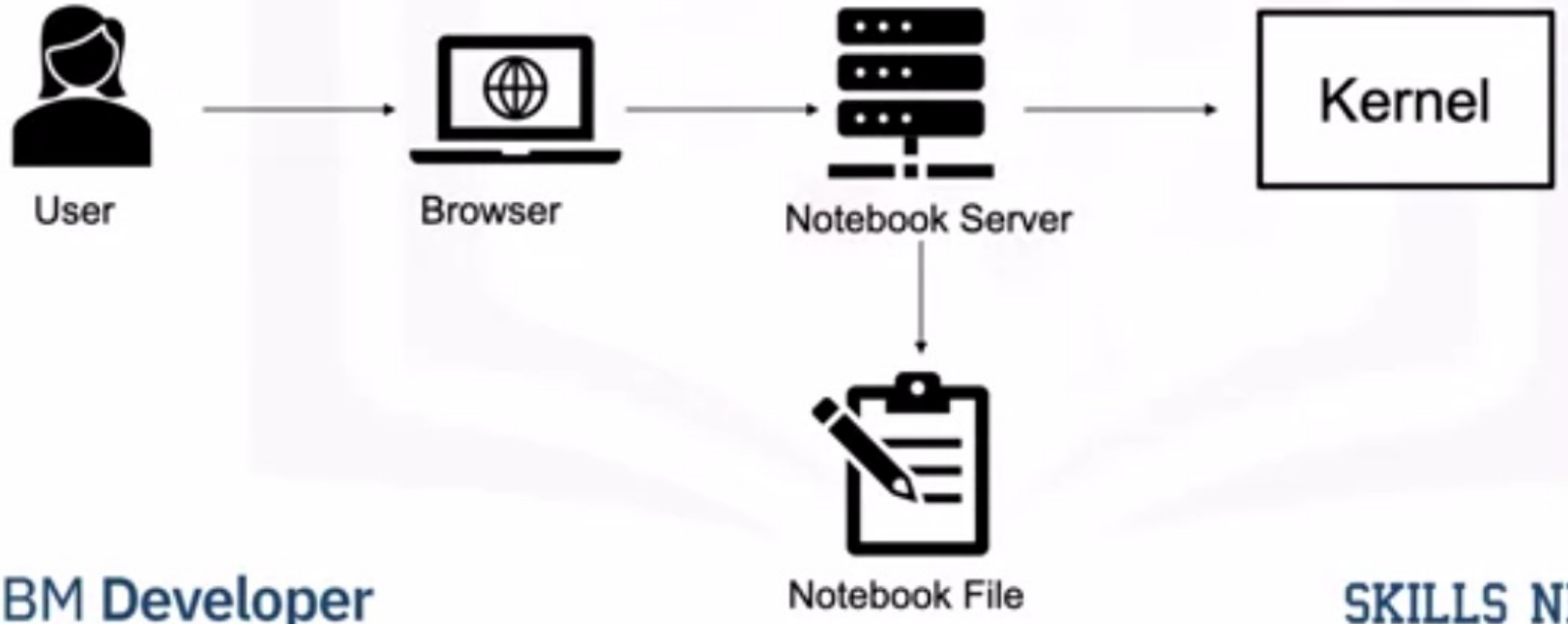
- A notebook kernel is a computational engine that executes the code contained in a Notebook file
- Jupyter Kernels for many other languages exist
- When the notebook is executed , the kernel performs the computation and produces the results.

Architecture

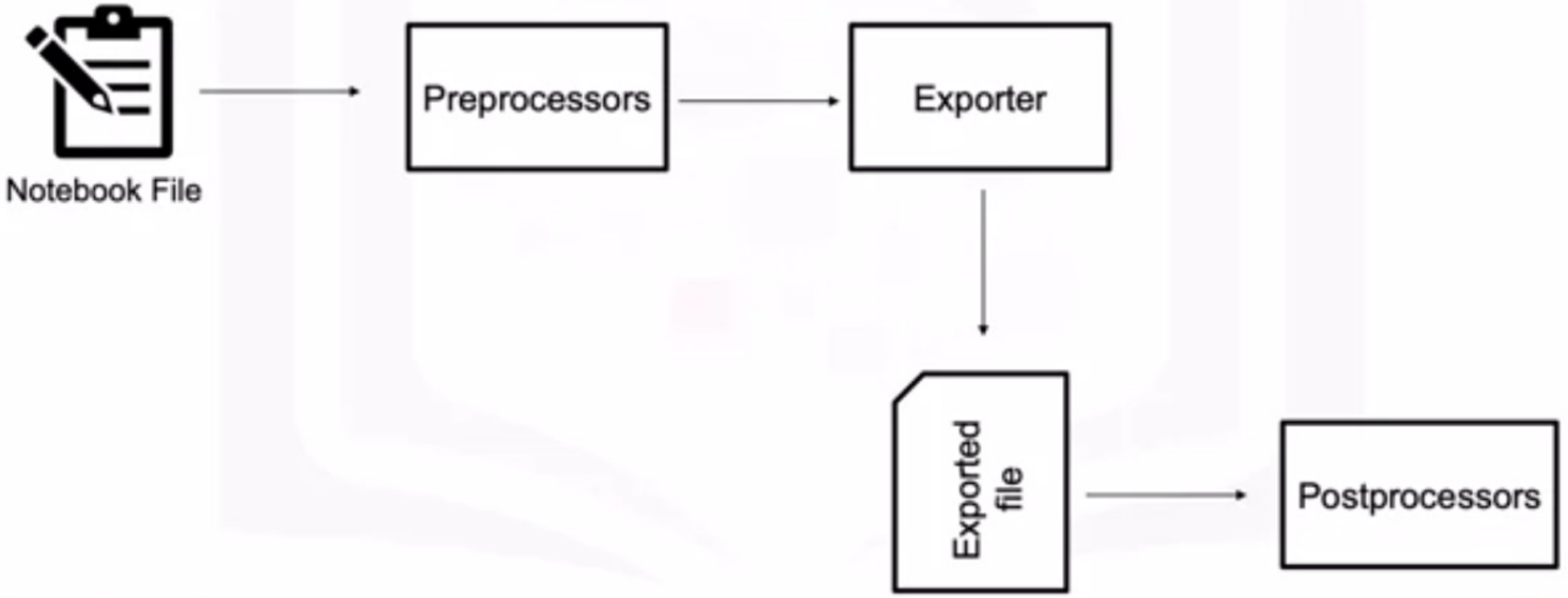
- Jupyter implements a two-process model, with a kernel and a client
- The client is the interface offering the user the ability to send code to the kernel
- The kernel executes the code and returns the result to the client for display
- The client is the browser when using a Jupyter notebook

Architecture

- Jupyter notebooks is used to represent code, metadata, contents and outputs
- It has a flexible interface that extends beyond code and output



Architecture



- [x] All of the above.

2. Which statement is true about Jupyter Notebook? **1 point**

- [ ] Jupyter Notebook is a commercial product of IBM.
- [x] Jupyter Notebook is free and open source.

3. What is a Jupyter Notebook kernel? **1 point**

- [ ] It is part of the operating system the Jupyter server runs on.
- [x] It is a wrapper running on the Jupyter server encapsulating the programming language interpreter.

Coursera Honor Code Learn more

- [ ] I, **Tushar Raha**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit | Save draft

R Language

What is R?

- Statistical programming language
- Used for data processing and manipulation
- Statistical, Data analysis and Machine learning
- R is used most by academics, healthcare and the government.
- R supports importing data from different sources: Flat files, Databases, Web, Statistical software, etc

R Capabilities

- It is easy to use compared to other Data Science tools
- Great tool for Visualization
- Basic Data Analysis doesn't require installing packages

What is RStudio

- RStudio is an Integrated Development Environment (IDE).
- It increases productivity in running **R** programming language.

Popular R Libraries for Data Science

`dplyr`	Data Manipulation
`stringr`	String Manipulation
`ggplot`	Data Visualization
`caret`	Machine Learning

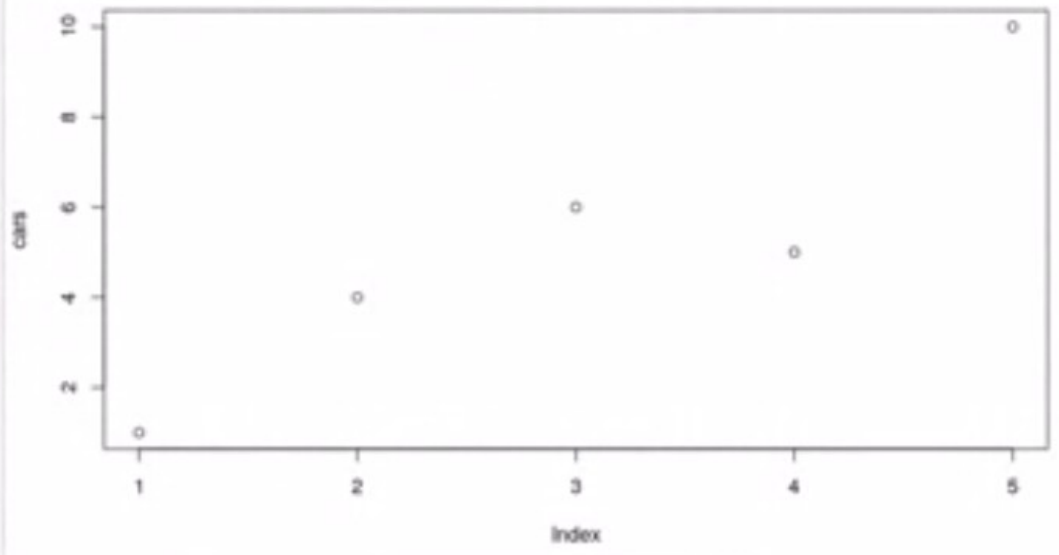
Data Visualization in R

R is a very great tool for data visualization and has different packages. Some of the popular and top data visualization are:

- ggplot – for data visualizations such as histograms, bar charts, scatterplots etc. It allows adding layers and components on a single visualization.
- Plotly - an R package can be used to create web-based data visualizations that can be displayed or saved as individual HTML files.
- Lattice - is a data visualization tool that is used to implement complex, multi-variable data sets.
- Leaflet – popular for creating interactive plots
- To install, use the command; install.packages("package name")

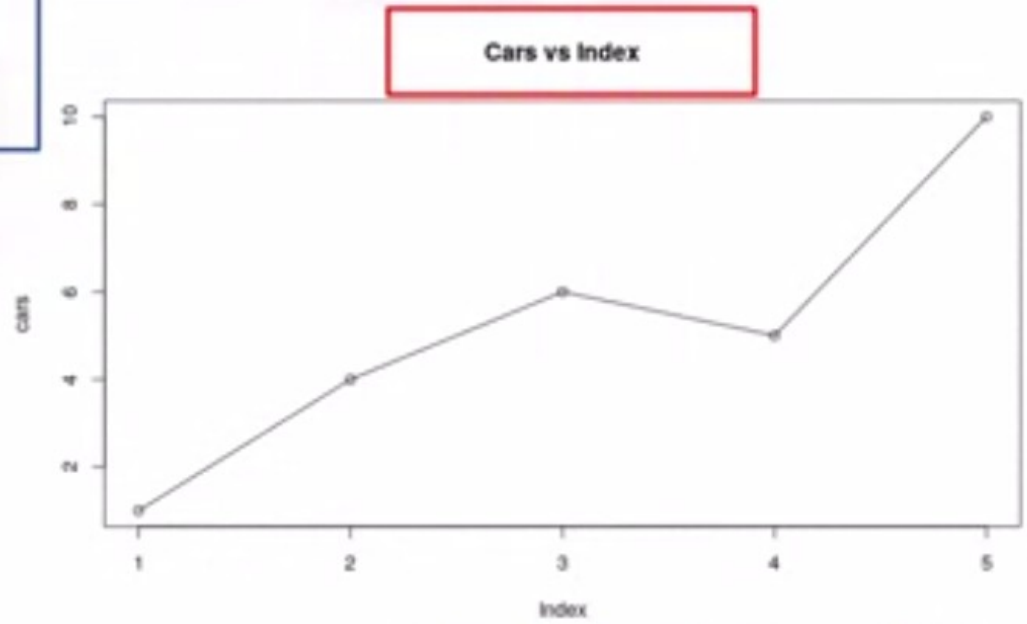
Using the plot function

```
# Define the cars vector with 5 values
cars <- c(1, 4, 6, 5, 10)
# Graph the cars vector with all defaults
plot(cars)
```



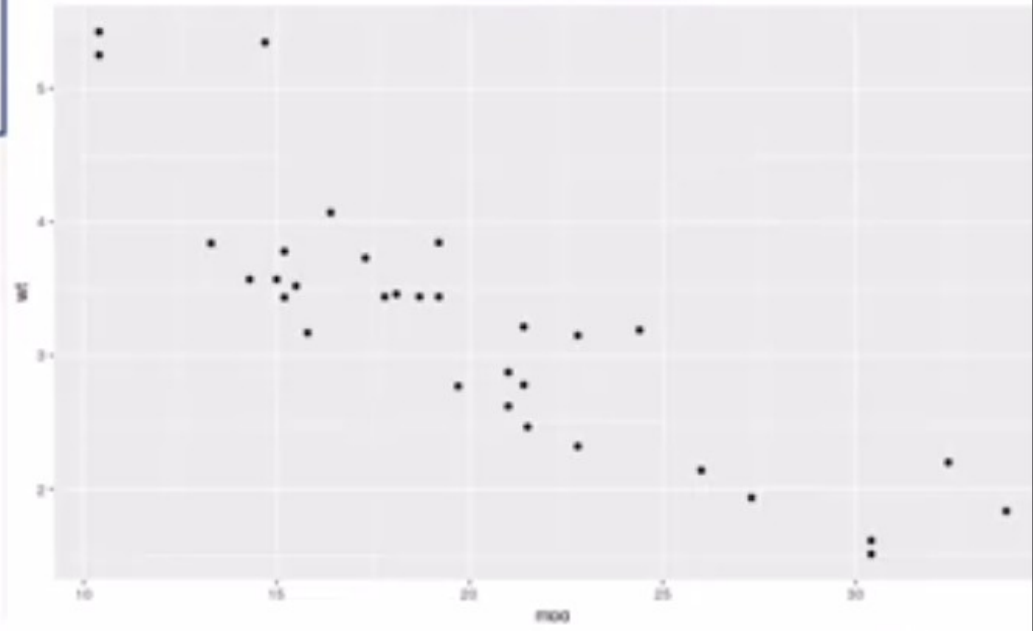
Using the plot function

```
# Define the cars vector with 5 values
cars <- c(1, 4, 6, 5, 10)
# Graph the cars vector with all defaults
plot(cars, type = "o")
# Create a title
title(main="Cars vs Index")
```

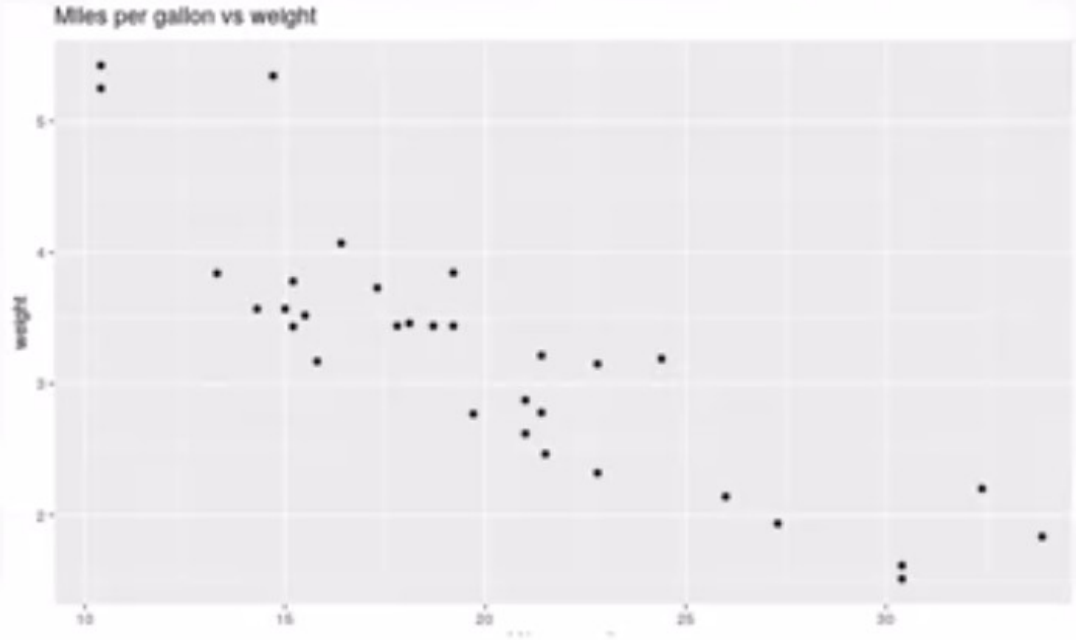


Using ggplot

```
library(ggplot2)
ggplot(mtcars, aes(x=mpg, y = wt))+geom_point()
```



```
ggplot(mtcars, aes(x=mpg, y = wt))+geom_point() + ggtitle("Miles per gallon vs weight") + labs(y="weight", x = "Miles per gallon")
```



Back

Practice Quiz - RStudio IDE

Practice Quiz • 9 min

Total points 3

1. Which of the following functions does RStudio unify? (Select all that apply.) **1 / 1 point**

- [ ] Storing of data.
- [x] Editing and execution of source code.

  Correct

- [x] Display of the R Console.

  Correct

- [x] Visualization of plots.

  Correct

- [x] Visualization of data in table form.

  Correct

GIT

Git

- Free and open source software
- Distributed version control system
- Accessible anywhere in the world
- One of the most common version control systems available
- Can also version control images, documents, etc.

git + GitHub

SHORT Glossary of Terms

SSH protocol – A method for secure remote login from one computer to another.

Repository - The folders of your project that are set up for version control.

Fork - A copy of a repository.

Pull request – The process you use to request that someone reviews and approves your changes before they become final.

Working directory – A directory on your file system, including its files and subdirectories, that is associated with a git repository.

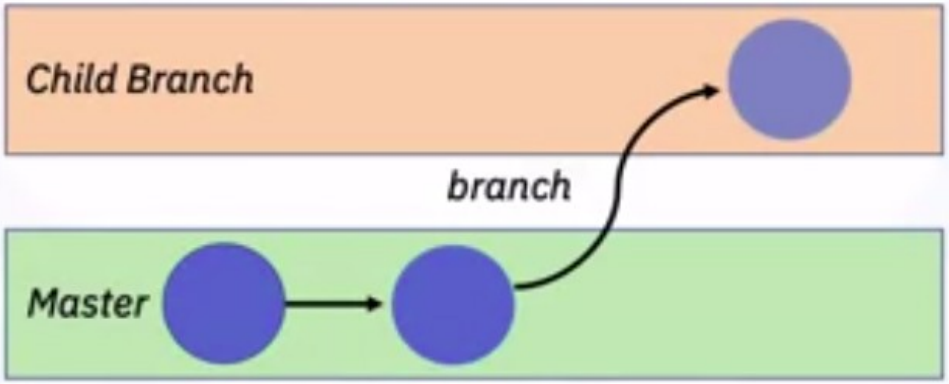
Basic Git Commands

- init
- add
- status
- commit
- reset
- log
- branch
- checkout
- merge

Creating a Branch – What?

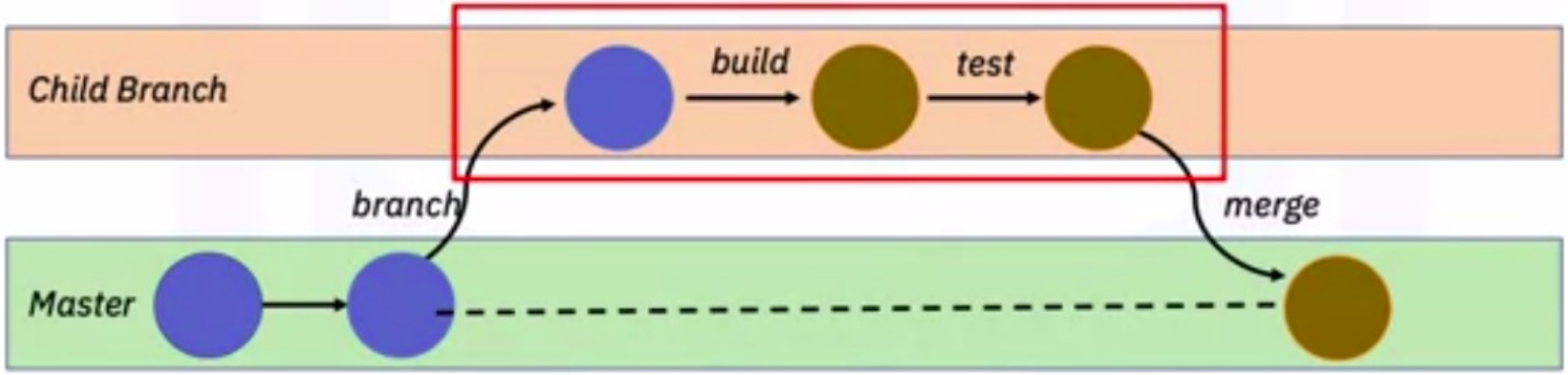
A branch is a snapshot of your repository.

- Master Branch is the official version of the project
- The child branch creates a copy of the master branch



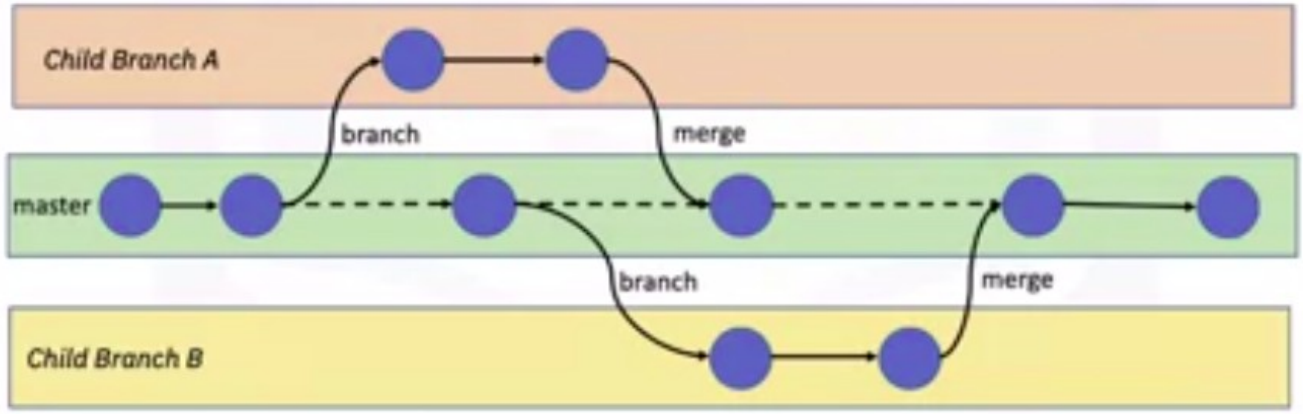
Creating a Branch – Why?

- Edits and changes are made in the child branch
- Tests are done to ensure quality before merging to the Master Branch



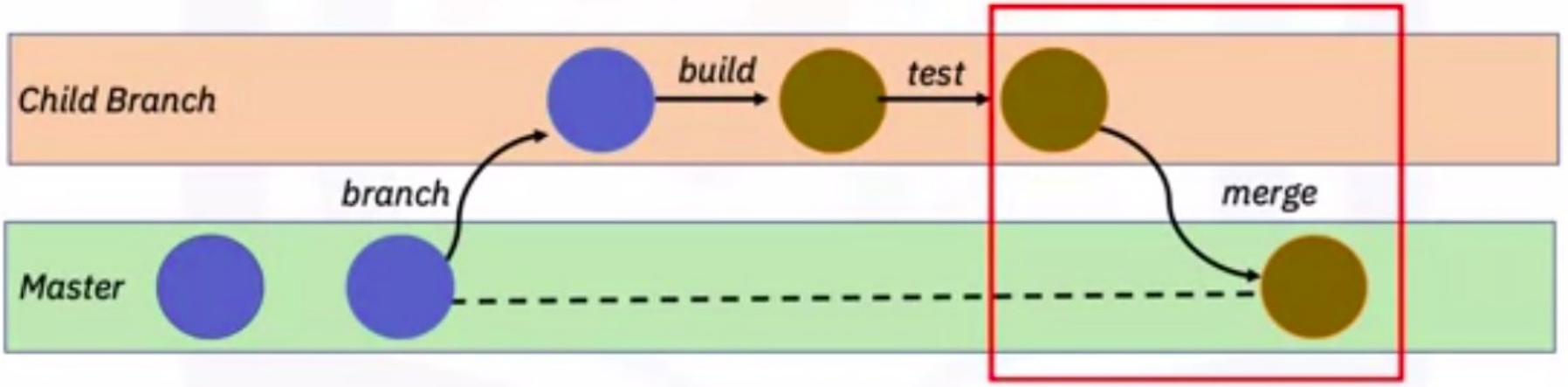
Creating a Branch – Why?

- Branches allow for simultaneous development and testing by multiple team members.



Pull Request (PR)

- Pull requests are a way of proposing changes to the main branch
- Ideally, another team member reviews the changes and approves it to be merged to the Master branch



Congratulations! You passed!

Grade received 100% **To pass** 70% or higher

Go to next item

Practice Quiz - GitHub

Total points 3

1. Which of the following statements are true? (Select all that apply.) **1 / 1 point**

- [x] Git is a system for version control of source code.

  Correct

- [ ] Git is an integrated development environment for data science.
- [x] Git is very useful for data science as well, since data science often involves a lot of source code to be written and managed.

  Correct

2. Which of the following statements about repositories are correct? (Select all that apply.) **1 / 1 point**

- [ ] The remote repository is only accessible by myself.
- [x] The local repository is only accessible by myself.

Correct

- [x] The staging is only accessible by myself.

Correct

- [x] The remote repository is accessible by all contributors.

Correct

- [ ] The local repository is accessible by all contributors.

3. What is the best process for contributing a bugfix to a foreign repository? **1 / 1 point**

- ( ) Ask the repository owner for write access to the repository.
- ( ) Send the fix via email to the author.